



TITLE:

Computational Methods for Analyzing  
Protein Complexes and Protein-Protein  
Interactions( Abstract\_要旨 )

AUTHOR(S):

Ruan, Peiying

---

CITATION:

Ruan, Peiying. Computational Methods for Analyzing Protein Complexes and Protein-Protein Interactions. 京都大学, 2015, 博士(情報学)

ISSUE DATE:

2015-03-23

URL:

<https://doi.org/10.14989/doctor.k19109>

RIGHT:

( 続紙 1 )

京都大学	博士 ( 情報 学 )	氏名	阮 佩穎 (Peiying Ruan)
論文題目	Computational Methods for Analyzing Protein Complexes and Protein-Protein Interactions タンパク質複合体および相互作用の情報解析手法		
(論文内容の要旨)			
<p>本論文は、タンパク質相互作用データからのタンパク質複合体の予測手法、および、それと関連するタンパク質のドメイン組成データの圧縮手法、遺伝子発現データからの遺伝子選択手法について述べられており、7章から構成されている。</p> <p>第1章では、研究の背景と動機、関連研究、対象とする問題の定義について概説するとともに、本論文で提案する解析手法の概要と論文の構成について述べている。</p> <p>第2章では、研究の基盤となっているサポートベクターマシン (SVM)、関連ベクターマシン (RVM)、カーネル関数、(重み付き) タンパク質相互作用ネットワーク、最小全域木、文法圧縮について説明している。</p> <p>第3章では、タンパク質の進化に関するより深い理解を得ることを主目的として、指定された生物種のすべてのタンパク質のドメイン組成データを文法圧縮により圧縮する問題を定義し、それが有向ハイパーグラフにおける一種の最小全域木を計算する問題に帰着されることを示している。この問題に対して、整数計画法に基づく計算手法を提案するとともに、ヒューリスティックの導入による高速化手法も提案している。そして複数の生物種のタンパク質ドメイン組成データを用いた計算機実験により、WWWデータに対する既存の圧縮手法をタンパク質に適用した場合よりも、提案手法がより効率的な圧縮が可能であることを示している。さらに、この結果から遺伝子融合のタンパク進化に果たす役割の重要性を示唆している。</p> <p>第4章では、遺伝子発現データなどを用いた遺伝子選択問題を対象としている。なお、遺伝子の多くはタンパク質に翻訳されることによりその機能を発現するので、遺伝子選択問題は重要なタンパク質の選択に役立つ可能性がある。この遺伝子選択のために近傍データをもとに重みを逐次更新していくRELIEFという計算手法が知られていた。この章ではこのRELIEFの考え方を発展させ、近傍データをもとに構成される局所超平面という局所的な情報をもとに重みを逐次更新していくことにより遺伝子選択を行う手法LHRを提案している。そして、人工データおよび複数の種類の遺伝子発現データを用いた計算機実験を行い、LHRが既存手法と比較してより優れた選択能力を持つことを示している。</p> <p>第5章では、重み付きのタンパク質相互作用ネットワークデータから、2個のタンパク質からなるタンパク質複合体を予測する問題を取り扱っている。この問題に対して、相互作用ネットワークに基づく5種類の特徴量とドメイン組成に基づく2種類の特徴量を定義し、さらに、ドメイン組成に基づくカーネル関数を定義した上で、それらをSVMと組み合わせた予測手法を提案している。そして、実際のタンパク質相互作用ネットワークデータおよびタンパク質複合体データを用いた計算機実験を行い、既存手法と比較してはるかに優れた予測率が得られることを示している。</p> <p>第6章では、重み付きのタンパク質相互作用ネットワークデータから、3個のタンパク質からなるタンパク質複合体を予測する問題を取り扱っている。第5章で用いた特徴量と類似の特徴量とカーネル関数を定義し、それとSVMを組み合わせた予測手法を提案するとともに、その予測結果をまとめたものをさらに特徴量として利用して予測を行う二段階の予測手法を提案している。なお、二段目の予測器としてSVMを用いる方法、および、RVMを用いる方法の二種類を提案している。そして、実際のタンパク質相互作用ネットワークデータおよびタンパク質複合体データを用いた計算機実験を行い、提案手法の中でSVMを二段目にも用いる二段階の予測手法が最も高精度の予</p>			

測率を持つことを示すとともに、それが既存手法と比較しても優れていることを示している。

第7章は結論であり、本研究をまとめるとともに、今後の研究の方向性や課題について述べている。

注) 論文内容の要旨と論文審査の結果の要旨は1頁を38字×36行で作成し、合わせて、3,000字を標準とすること。

論文内容の要旨を英語で記入する場合は、400～1,100 wordsで作成し  
審査結果の要旨は日本語500～2,000字程度で作成すること。

(論文審査の結果の要旨)

本論文は、タンパク質複合体の予測手法と、これに深く関連する問題である、タンパク質ドメイン構成データの解析手法と遺伝子発現データを用いた遺伝子選択手法について述べたものであり、得られた成果は以下のとおりである。

(1) 指定された生物種のすべてのタンパク質のドメイン組成データを文法圧縮の考え方に基づき圧縮するという問題を定義し、それがハイパーグラフにおける一種の最小全域木を計算する問題に帰着されることを示した。そして、この問題に対する整数計画法に基づく計算手法と、それとヒューリスティックを組み合わせた計算手法を開発した。複数の生物種に対するタンパク質ドメイン組成データを用いた計算機実験によりその有効性を示した。

(2) 遺伝子発現データを用いて重要な遺伝子を選択することの定式化である遺伝子選択問題に対して、近傍データをもとに構成される局所超平面を用いて重みを逐次更新していくことにより選択を行う計算手法を開発した。そして、人工データおよび複数の種類の遺伝子発現データを用いた計算機実験を行い、開発した手法が既存手法と比較して優れた性能を持つことを示した。

(3) 重み付きタンパク質相互作用ネットワークデータから2個のタンパク質よりなるタンパク質複合体を予測する問題に対して、ネットワークデータおよびタンパク質ドメイン組成データに基づく7種類の特徴量を定義し、さらにドメイン組成に基づくカーネル関数を定義し、それらとサポートベクターマシンを組み合わせた計算手法を開発した。そして、実際のタンパク質データを用いた計算機実験により、開発手法の既存手法に対する優位性を示した。

(4) 重み付きタンパク質相互作用ネットワークデータから3個のタンパク質よりなるタンパク質複合体を予測する問題に対して、(3)におけるのと同様の予測方式を開発するとともに、その予測結果を特徴量として用いて予測を行う二段階の予測方式を開発した。そして、実際のタンパク質データを用いた計算機実験により、二段階手法の一段階手法に対する優位性、および、二段階手法の既存手法に対する優位性を示した。

以上、本論文ではバイオインフォマティクスにおける重要な研究課題であるタンパク質ドメイン組成データの情報解析、遺伝子発現データからの遺伝子選択、タンパク質相互作用ネットワークからの複合体予測のそれぞれに対して新規で有用な計算手法を提案するとともに、実データを用いた計算機実験により、それらの有効性を示している。提案手法のいずれもが新規性、有用性、拡張性の高いものであり、当該分野の発展のために十分な寄与をしている。よって、本論文は博士(情報学)の学位論文として価値あるものと認める。また、平成27年2月23日に実施した論文内容とそれに関連した口頭試問を行った結果合格と認めた。

注) 論文審査の結果の要旨の結句には、学位論文の審査についての認定を明記すること。  
更に、試問の結果の要旨(例えば「平成 年 月 日論文内容とそれに関連した口頭試問を行った結果合格と認めた。」)を付け加えること。

Webでの即日公開を希望しない場合は、以下に公開可能とする日付を記入すること。

要旨公開可能日： 年 月 日以降